Nonparametric inference for general categorical time series with time-varying parameters

Lionel Truquet (CREST-ENSAI, Rennes),

Outline

1. Categorical time series and locally stationary processes
2. Locally stationary categorical time series and Markov chains in random environments (MCRE)
3. An approach using iterated random maps systems

Categorical time series

- Encountered in various contexts: binary (failures in industrial process, trading activity, growth/recession periods in economics) or multinomial (air/water quality, weather types, DNA) or multiple binary (absence/presence of species in ecology, dynamic random graphs for social networks).
- There exist some stationary models, e.g Dynamic probit model ([Kaufmann (1987), AoS], [de Jong and Woutersen (2011), ET], [Kauppi and Saikkonen (2008), Rev. Econ. Stat.]).
- Multinomial or ordinal autoregressive time series models also exist ([Fokianos and Kedem (2003), St. Sc.], [Fokianos and Moysiadis (2014,2017), JMVA]).
- **Lack of theory**: ergodicity, mixing properties, limit theorems, extension to time-varying parameters, non-stationary models. [Fokianos and T. (2019), SPA], [T. (2020), Bernoulli] provide a few elements.

Example of non-stationarity in finance

Binary time series[Fokianos and Moysiadis, 2017, JMVA]

Trading activity of a thinly traded share at the Johannesburg Stock Exchange from 5th of October 1987 to 3rd of June 1991. The data are binary, with a value equal to 1 if a trade has been recorded at time t and 0 otherwise.

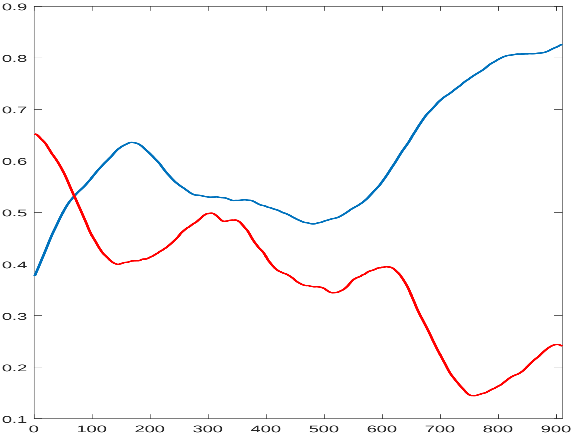

Figure: Local estimation of the probability of 1 using a sliding window

Context of local stationarity (1)

- Locally stationary models have been mainly developed for continuous state spaces, in the linear and non-linear case [Dahlhaus et al. (2019), Bernoulli].

- Examples include infinite-moving averages $X_{n,t} = \sum_{j\geq 0} \theta_j(t/n)\varepsilon_{t-j}$ [Dahlhaus (1997), AoS] or other forms of non-linear processes, e.g. ARCH processes

$$X_{n,t} = \varepsilon_t \sqrt{\theta_0(t/n) + \sum_{j\geq 1} \theta_j(t/n) X_{n,t-j}^2}.$$

See [Dahlhaus and Subba Rao (2006), AoS], [Fryzlewicz et al. (2008), AoS], [T. (2016), JRSSB].

- A general theory for local inference of time-varying parameters $u \mapsto \theta(u) = (\theta_j(u))_j$ is based on the $\mathbb{L}^p-$approximations

$$\|X_{n,t} - X_t(u)\|_p \leq C\left[|u - t/n| + 1/n\right], \quad \|X_t(u) - X_t(v)\|_p \leq C|u - v|.$$

where $(X_t(u))_{t\in\mathbb{Z}}$ denotes a stationary processes with fixed parameters $\theta(u)$.

Context of local stationarity (2)

- A natural inference method for parameter curves is to localize log-likelihood methods, i.e.

$$\hat{\theta}(u) = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} w_t(u) \ell_{n,t}(\theta),$$

where

$$w_t(u) = K\left(\frac{u - t/n}{b}\right) / \sum_{j=1}^{n} K\left(\frac{u - j/n}{b}\right).$$

- Asymptotic properties requires to control the behavior of some localized partial sums

$$S_n = \sum_{t=1}^{n} w_t(u) f\left(X_{n,t}, X_{n,t-1}, \dots\right).$$

- Typically

$$\begin{aligned} S_n - \mathbb{E} f\left(X_{t-j}(u); j \geq 0\right) &= S_n - \mathbb{E}S_n + \mathbb{E}S_n - \mathbb{E} f\left(X_{t-j}(u); j \geq 0\right) \\ &= \text{Variance part} + \text{Bias part}. \end{aligned}$$

Context of local stationarity (3)

- The variance part can be controlled using dependence properties (e.g. mixing conditions).
- To control the bias part, we only need to control local approximations of expectations.
- The theory developed for continuous state-spaces can be generalized if we replace local approximations of the state process by local approximations of expectations. See [T. (2019), AoS], [T. (2020), Bernoulli] for a study of locally stationary Markov chains.

Locally stationary categorical time series

- Our aim is to study locally stationary time series defined by

$$\begin{aligned}\mathbb{P}\left(Y_{n,t}=y|Y_{n,t-1},X_{n,t-1},\ldots\right) &= \mu\left(\theta(t/n),Y_{n,t-1},X_{n,t-1};y\right)\\ &:= P_{n,t}\left(Y_{n,t-1},y\right)\end{aligned}$$

where $(X_{n,t})_{1\leq t\leq n}$ is a locally stationary process in the usual sense (with $\{X_t(u),(t,u)\in\mathbb{Z}\times[0,1]\}$ as stationary approximations). $y\in E$ **finite**.

- This framework covers the case of binary time series

$$P_{n,t}\left(Y_{n,t-1},y\right)=F\left[\theta_0(t/n)+\theta_1(t/n)Y_{n,t-1}+\theta_2(t/n)X_{n,t-1}\right],$$

with logistic or probit models (see [Fokianos and Kedem (2003), Stat. Sc.] for stationary versions).

- More complicated models (multinomial, ordinal) categorical time series can be defined in this way.

- One can also include other lags $Y_{n,t-2},\ldots$ but taking vector components, one can go back to a dynamic of order 1.

Outline

1. Categorical time series and locally stationary processes
2. Locally stationary categorical time series and Markov chains in random environments (MCRE)
3. An approach using iterated random maps systems

General notations

- We first assume strict exogeneity [Chamberlain (1982), Econometrica], i.e. $Y_{n,t}$ and $X_{n,t}, X_{n,t+1}, \ldots$ are independent conditionally on $(X_{n,t-j}, Y_{n,t-j})_{j\geq 1}$.
- This exogeneity notion is restrictive but easier to handle. Set

$$\begin{aligned} P_{n,t}(y,y') &= \mathbb{P}\left(Y_{n,t} = y' | Y_{n,t-1} = y, X_{n,t-1}\right), \\ P_t(u)(y,y') &= \mathbb{P}\left(Y_t(u) = y' | Y_{t-1}(u) = y, X_{t-1}(u)\right). \end{aligned}$$

- Conditionally on the covariates path, the dynamic is then given by a time-inhomogeneous Markov chains with the previous random stochastic matrices.
- Stationary solutions exist if the product $P_1(u)\cdots P_m(u)$ has positive entries a.s. for some m [Kifer (1995), Math Z]. The principle is to define

$$\pi_t(u)_y = \mathbb{P}\left(Y_t(u) = y | X_{t-j}(u), j \geq 1\right) = \lim_{N\to\infty} P_{t-N}(u)\cdots P_{t-1}(u)(z,y).$$

We have the equilibrium relations $\pi_t(u)P_t(u) = \pi_{t+1}(u)$ a.s.

Approximation result (1)

Set for $t \leq 0$, $X_{n,t} = X_t(0)$ and $P_{n,t} = P_t(0)$.

A1 There exist $r > 1$ and $C > 0$ such that for all $u, v \in [0, 1]$ and $1 \leq t \leq n$,

$$\|P_{n,t} - P_t(u)\|_r \leq C\left[|u - t/n| + 1/n\right], \quad \|P_t(u) - P_t(v)\|_r \leq C|u - v|.$$

A2 The mixing coefficients of $(X_t(u))_{t \in \mathbb{Z}}$ satisfy

$$\alpha_{X(u)}(k) \leq L\,(1 + k)^{-\rho}, \quad k \in \mathbb{N},$$

for some $\rho > 2$.

Approximation result (2)

If $M > 0$, let $\mathcal{L}_M$ the set of measurable mappings $f : \mathbb{R}^d \times E^2 \to \mathbb{R}$ such that there exists a positive constant L such that for all (x,z) and $(x',z') \in \mathbb{R}^d \times E^2$,

$$|f(x,z) - f(x',z')| \leq L\left(1 + |x|^M + |x'|^M\right)\left(|x - x'| + \mathbb{1}_{z \neq z'}\right).$$

Theorem

Suppose that Assumptions **A1-A2** *hold true. Let* $1 \leq q < \rho/2$ *and* $1/p = 1/q + 1/r$. *Suppose furthermore that* $f \in \mathcal{L}_M$ *and that for* $s = (M+1)p/(p-1)$,

$$\sup_{n \geq 1} \max_{1 \leq t \leq n} \|X_{n,t}\|_s < \infty, \qquad \sup_{u \in [0,1]} \|X_0(u)\|_s < \infty.$$

There then exists $\widetilde{C} > 0$ *such that for all* $n \geq 1$, $1 \leq t \leq n$ *and* $(u,v) \in [0,1]^2$,

$$\left|\mathbb{E} f\left(X_{n,t-1}, Y_{n,t-1}, Y_{n,t}\right) - \mathbb{E} f\left(X_{t-1}(u), Y_{t-1}(u), Y_t(u)\right)\right| \leq \widetilde{C}\left[|u - t/n| + 1/n\right],$$

$$\left|\mathbb{E} f\left(X_{t-1}(u), Y_{t-1}(u), Y_t(u)\right) - \mathbb{E} f\left(X_{t-1}(v), Y_{t-1}(v), Y_t(v)\right)\right| \leq \widetilde{C}\,|u - v|.$$

Derivatives and bias control

- Our aim here is to control the regularity of expectations of some functionals of the stationary approximations $(X_t(u), Y_t(u))_{t\in\mathbb{Z}}$.
- For continuous state spaces, there exists in the literature a notion of derivative process ($u \mapsto X_t(u)$ has derivative $u \mapsto \dot{X}_t(u)$) either in a a.s. sense or at least in some $\mathbb{L}^r-$norms.
- Example. $X_t(u) = \varepsilon_t - \theta(u)\varepsilon_{t-1}$ gives $\dot{X}_t(u) = \varepsilon_t - \dot{\theta}(u)\varepsilon_{t-1}$.
- The derivative of $u \mapsto \mathbb{E}f(X_0(u))$ is given by $\mathbb{E}f'(X_0(u))\dot{X}_0(u)$ under some smoothness and integrability conditions.
- As usual in nonparametric curve estimation, existence of such derivatives allows to control the bias of localized partial sums (and to get a smaller bias).

Perturbation methods for random measures (1)

- For categorical processes, the notion of derivative for $u \mapsto Y_t(u)$ is not adapted.

- However, one can write for instance

$$\mathbb{E}f\left(X_{t-1}(u), Y_t(u)\right) = \sum_{y \in E} \mathbb{E}f\left(X_{t-1}(u), y\right) \pi_t(u)(y).$$

- If $u \mapsto X_t(u)$ admits a derivative, one can try to prove existence of a derivative for $u \mapsto \pi_t(u)$ (which depends on $X_{t-j}(u)$, $j \geq 1$).

- The equilibrium relations $\pi_t(u)P_t(u) = \pi_{t+1}(u)$ "should" give a derivative process $\left(\dot{\pi}_t(u)\right)_{t \in \mathbb{Z}}$ defined by

$$\dot{\pi}_{t+1}(u) = \dot{\pi}_t(u)P_t(u) + \pi_t(u)\dot{P}_t(u).$$

Perturbation methods for random measures (2)

Theorem

Suppose that the previous assumptions hold true and that $u \mapsto P_t(u)$ is (continuously) differentiable in $\mathbb{L}^r-$norm. The mapping $u \mapsto \pi_t(u)$ is also (continuously) differentiable in another $\mathbb{L}^p-$norm with $p \leq r$.

- One can also obtain a result for the differentiability of

$$u \mapsto \mathbb{P}\left(Y_{t-1}(u) = y, Y_t(u) = y' | X_{\cdot}(u)\right) = \pi_{t-1}(u)(y) P_t(u)(y, y').$$

- Differentiability of $u \mapsto \mathbb{E} f\left(X_{t-1}(u), Y_t(u)\right)$ or $u \mapsto \mathbb{E} f\left(X_{t-1}(u), Y_{t-1}(u), Y_t(u)\right)$ can then be obtained for differentiable mappings f w.r.t. its first argument (under some moment and growth conditions).

- Under additional conditions, higher-order derivatives can be obtained.

Mixing properties for the variance part

Strong mixing properties of the pair $((X_{n,t}, Y_{n,t}))_{1\le t\le n}$.
The uniform $\alpha-$mixing coefficients of a triangular array $(X_{n,t})_{1\le t\le n}$:

$$\alpha_X^{(n_0)}(k) = \sup_{n\ge n_0\ge k} \max_{1\le t\le n-k} \alpha\left(\mathcal{F}_{n,t}, \mathcal{G}_{n,t+k}\right),$$

where

$$\mathcal{F}_{n,t} = \sigma\left(X_{n,s} : 1 \le s \le t\right), \quad \mathcal{G}_{n,t} = \sigma\left(X_{n,s} : t \le s \le n\right).$$

In what follows, we set $V_{n,t} = (X_{n,t}, Y_{n,t})$ for $1 \le t \le n$.

Theorem

Suppose that the previous assumptions hold true.

1. *If* $\alpha_X^{(n_0)}(k) = O\left(\kappa^k\right)$ *for some* $\kappa \in (0,1)$, *there exists* $n_0' \ge n_0$ *and* $\overline{\kappa} \in (0,1)$ *s.t.* $\alpha_V^{(n_0')}(k) = O\left(\overline{\kappa}^{\sqrt{k}}\right)$.
2. *If* $\alpha_X^{(n_0)}(k) = O\left(k^{-\kappa}\right)$ *for some* $\kappa > 1$, *there exists* $n_0' \ge n_0$ *s.t. for* $0 < \kappa' < \kappa - 1$, $\alpha_V^{(n_0')}(k) = O\left(k^{-\kappa'}\right)$.

Consequences

- The previous result apply for local inference in the standard autoregressive categorical time series models (logistic, probit, multinomial...) with Lipschitz coefficients.
- Under the previous assumptions, the bias part of localized partial sums is of order b (up to a term of order $1/n$).
- $(nb)^{-1/2}\times$ the variance part is asymptotically Gaussian.
- Pointwise consistency and asymptotic normality of $u \mapsto \hat{\theta}(u)$ follows under standard conditions.

Outline

Sequential exogeneity. Motivation with a simple example

$$Y_{n,t} = \begin{cases} 1 \text{ if } \theta_0(t/n) + \theta_1(t/n)Y_{n,t-1} + \theta_2(t/n)X_{n,t-1} + \varepsilon_t > 0, \\ 0 \text{ otherwise.} \end{cases}$$

- We recover logistic or probit models according to the noise distribution, **but the previous results about MCRE only apply when the two processes $(X_{n,t})_t$ and $(\varepsilon_t)_t$ are independent due to strict exogeneity**.
- Strict exogeneity does not allow an impact of ε_t on $X_{n,t}, X_{n,t+1}, \ldots$.
- **Sequential exogeneity or predetermindness** is often more attractive. ε_t is independent from $\sigma\left((X_{n,t-j}, \varepsilon_{t-j}) : j \geq 1\right)$.

Locally stationary iterated random maps system

$$Y_{n,t} = F\left(\theta(t/n), Y_{n,t-1}, X_{n,t-1}, \varepsilon_t\right) := f_{n,t}\left(Y_{n,t-1}\right),$$
$$Y_t(u) = F\left(\theta(u), Y_{t-1}(u), X_{t-1}(u), \varepsilon_t\right) := f_t(u)\left(Y_{t-1}(u)\right).$$

S1 For every $u \in [0,1]$, the process $(f_t(u))_{t\in\mathbb{Z}}$ is stationary and ergodic.

S2 There exist an integer $m \geq 1$ and a real number $\kappa > 0$ such that for all $u \in [0,1]$,

$$\mathbb{P}\left(\# f_t(u) \circ \cdots \circ f_{t-m+1}(u)(E) = 1\right) \geq \kappa.$$

S3 There exist two positive real numbers L and ρ such that for all $u \in [0,1]$,

$$\alpha_{f_{.}(u)}(k) \leq L(1+k)^{-\rho}, \quad k \in \mathbb{N}.$$

S4 There exists a positive constant C such that for any $(n,u,v) \in \mathbb{N}^* \times [0,1]^2$ and $1 \leq t \leq n$,

$$\mathbb{P}\left(f_{n,t} \neq f_t(u)\right) \leq C\left[|u - t/n| + 1/n\right],$$

$$\mathbb{P}\left(f_t(u) \neq f_t(v)\right) \leq C|u-v|.$$

Local approximation result (1)

Theorem

For any real number p such that $1 < p < (1+\rho)/2$, there exists $C_0 > 0$ s.t. for all $n \geq 1$ and $(u, v) \in [0, 1]$,

$$\mathbb{P}\left(Y_{n,t} \neq Y_t(u)\right) \leq C_0 \left[\left|u - \frac{t}{n}\right| + \frac{1}{n}\right]^{\frac{p-1}{p}}, \quad \mathbb{P}\left(Y_t(u) \neq Y_t(v)\right) \leq C_0 \left|u - v\right|^{\frac{p-1}{p}}.$$

When the mixing rate is exponential, $p = \infty$ and we recover the usual approximation bounds for locally stationary processes.

Local approximation result (2)

S5 $\sup_{n\geq 1, 1\leq t\leq n} \|X_{n,t}\|_q, \sup_{u\in[0,1]} \|X_1(u)\|_q < \infty$ for $q > M$,

S6 $\|X_{n,t} - X_t(u)\|_{\frac{q}{q-M}} \leq C_1 \left[\left|u - \frac{t}{n}\right| + \frac{1}{n}\right]$,

$\|X_t(u) - X_t(v)\|_{\frac{q}{q-M}} \leq C_1 |u - v|$.

Theorem

There exists $\widetilde{C} > 0$ s.t. for all $n \geq 1$, $(u, v) \in [0,1]^2$, $1 \leq t \leq n$ and $1 < p < (1+\rho)/2$,

$$\left|\mathbb{E}f\left(X_{n,t}, Y_{n,t-1}, Y_{n,t}\right) - \mathbb{E}f\left(X_t(u), Y_{t-1}(u), Y_t(u)\right)\right| \leq \widetilde{C}\left[|u - t/n| + \frac{1}{n}\right]^{\frac{(p-1)(q-M)}{pq}}$$

$$\left|\mathbb{E}f\left(X_t(u), Y_{t-1}(u), Y_t(u)\right) - \mathbb{E}f\left(X_t(v), Y_{t-1}(v), Y_t(v)\right)\right| \leq \widetilde{C}\,|u - v|^{\frac{(p-1)(q-M)}{pq}}.$$

1. If the covariates are bounded and geometrically mixing, take $q = p = \infty$ and we recover the standard bounds (but the approximation rate is slower in general).
2. The bias of localized partial sums is now of order $b^{\frac{(p-1)(q-M)}{pq}}$ even if Lipschitz properties are assumed for the covariates.

Mixing conditions

The uniform $\alpha-$mixing coefficients of a triangular array $(X_{n,t})_{1\leq t\leq n}$:

$$\alpha_X^{(n_0)}(k) = \sup_{n\geq n_0\geq k} \max_{1\leq t\leq n-k} \alpha\left(\mathcal{F}_{n,t}, \mathcal{G}_{n,t+k}\right),$$

where

$$\mathcal{F}_{n,t} = \sigma\left(X_{n,s} : 1\leq s\leq t\right), \quad \mathcal{G}_{n,t} = \sigma\left(X_{n,s} : t\leq s\leq n\right).$$

In what follows, we set $V_{n,t} = (X_{n,t}, Y_{n,t})$ for $1\leq t\leq n$.

Theorem

Suppose that the previous assumptions hold true.

1. *If $\alpha_{X,\varepsilon}^{(n_0)}(k) = O\left(\kappa^k\right)$ for some $\kappa\in(0,1)$, there exists $n_0'\geq n_0$ and $\overline{\kappa}\in(0,1)$ s.t. $\alpha_V^{(n_0')}(k) = O\left(\overline{\kappa}^{\sqrt{k}}\right)$.*
2. *If $\alpha_{X,\varepsilon}^{(n_0)}(k) = O\left(k^{-\kappa}\right)$ for some $\kappa > 1$, there exists $n_0'\geq n_0$ s.t. for $0<\kappa'<\kappa-1$, $\alpha_V^{(n_0')}(k) = O\left(k^{-\kappa'}\right)$.*

Example. Ordinal time series on $E = \{0, 1, \ldots, N-1\}$ (1)

$$
\begin{aligned}
Y_{n,t} &= \sum_{k=0}^{N-1} k \mathbb{1}_{\{c_k(t/n, X_{n,t-1}) < Y^*_{n,t} < c_{k+1}(t/n, X_{n,t-1})\}}, \\
Y^*_{n,t} &= h_{t/n}\left(Y_{n,t-1}, \ldots, Y_{n,t-p}, X_{n,t-1}\right) + \varepsilon_t.
\end{aligned}
$$

E0 The process $((\varepsilon_t, X_t(u)))_{t \in \mathbb{Z}}$ is stationary and ergodic.

E1 For any $t \leq n$, ε_t is independent of $\mathcal{F}_{t-1}$ and has a Lipschitz cdf F_ε with a full support $\mathbb{R}$.

E2 $\exists C > 0$ s.t.

$$
\mathbb{E} \max_{y \in E^p} \left| h_{t/n}\left(y, X_{n,t-1}\right) - h_u\left(y, X_{t-1}(u)\right) \right| \leq C \left[|u - t/n| + 1/n \right],
$$

$$
\mathbb{E} \max_{y \in E^p} \left| h_u\left(y, X_{t-1}(u)\right) - h_v\left(y, X_{t-1}(v)\right) \right| \leq C \left| u - v \right|,
$$

$$
\mathbb{E} \max_{1 \leq k \leq N-1} \left| c_k\left(t/n, X_{n,t-1}\right) - c_k\left(u, X_{t-1}(u)\right) \right| \leq C \left[|u - t/n| + 1/n \right],
$$

$$
\mathbb{E} \max_{1 \leq k \leq N-1} \left| c_k\left(u, X_{t-1}(u)\right) - c_k\left(v, X_{t-1}(v)\right) \right| \leq C \left| u - v \right|.
$$

Example. Ordinal time series on $E = \{0, 1, \ldots, N-1\}$ (2)

$$\begin{aligned} Y_{n,t} &= \sum_{k=0}^{N-1} k \mathbb{1}_{\{c_k(t/n, X_{n,t-1}) < Y^*_{n,t} < c_{k+1}(t/n, X_{n,t-1})\}}, \\ Y^*_{n,t} &= h_{t/n}\left(Y_{n,t-1}, \ldots, Y_{n,t-p}, X_{n,t-1}\right) + \varepsilon_t. \end{aligned}$$

Proposition

E0-E1-E2 *imply* **S1-S2-S4**.

1. Under additional mixing assumptions on $(X_t(u), \varepsilon_t)_t$ or $(X_{n,t}, \varepsilon_t)_t$, one can control the variance of localized partial sums.
2. Note that the bias is of lower order in general. It is an open question to study differentiability properties with this second approach.

To take away

- It is possible to extend the theory of locally stationary processes for dealing with categorical time series.
- Many results can be obtained with two main notions of exogeneity.
- The key argument consists in changing the metric for either comparing the distributions of the categorical outcomes or looking at the Hamming distance.
- The more restrictive notion of strict exogeneity allows to get much sharper results.